

Two types of Hapaxomers

- Internal Palindromic Type II enzymes

– e.g., *Sfi* I

G G C C N N N[^]N G G C C

C C G G N[^]N N N C C G G

- Outside Cutters (Type IIS)

– e.g., *Sap* I

G C T C T C N[^]N N N

C G A G A A G N N N[^]

A

CAGNNCTG
CTCANNCTG

B

GGATGNNNNNNNNNNNN
CCTACNNNNNNNNNNNN
NNNNNNNNNNNNCATCC
NNNNNNNNNNNNGTAGG

Figs 2A-B

300

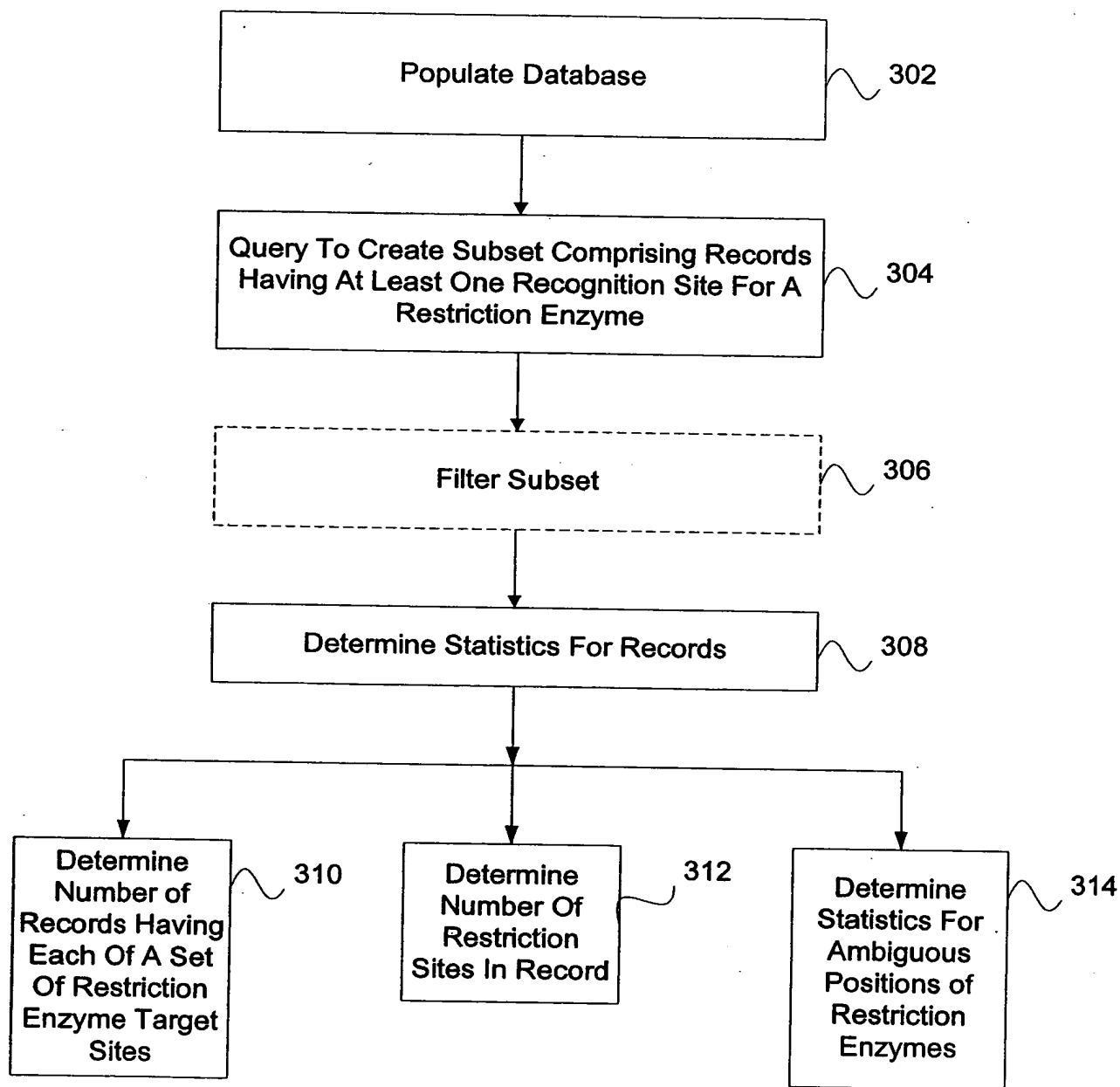


FIG. 3

	Sap I (0)	Sfl I (0)	Sfl I (0+1)	Sfl (0-2) Sgf I/Pme I	Sap I (0)	Sap I (0+1)	Sap I (0-2)
Hs_Fna	56.72	85.93	96.72	98.91	56.72	84.67	94.82
Mgc	61.98	87.02	97.94	99.70	61.98	90.31	97.60
E coll	86.71	99.35	100.00	100.00	86.71	98.37	99.53
C elegans	65.93	99.46	99.97	100.00	65.93	90.38	97.19
S cerevisiae	80.02	99.15	99.98	100.00	80.02	96.68	99.29
Arabidopsis	70.83	99.56	100.00	100.00	70.83	93.37	98.63

Fig 4

7+ Cutters

Enzymes	Recognition Sequence	HsFna	MGC	Ec	Ce	Sc	At
AarI	CACCTGCNNNN^NNNN	7142	5355				1138
Abel	CC^TCA GC not available	7970	5836	141	90	374	1833
AscI	GG^CGCG CC	515	336	152	10	13	26
AsiSI	GCG AT^CGC	108	62	207	39	29	178
BbvCI	CC^TCA GC	7970	5836	141	90	374	1833
CciNI	GC^GGCC GC	1444	823	19	33	31	97
CpoI	CG^GWC CG	1119	781	347			
CspI	CG^GWC CG	1119	781	347			
CspBI	GC^GGCC GC not available	1444	823	19	33	31	97
FseI	GG CCGG^CC	1139	740	5	9	10	70
MabI	A^CCWGG T						
MchAI	GC^GGCC GC not available	1444	823	19	33	31	97
Mlu1106I	RGGWCCY not available						
NotI	GC^GGCC GC	1444	823	19	33	31	97
PacI	TTA AT^TAA	708	395	66	8	213	138
Pfi27I	RG^GWC CY not available						
PpuMI	RG^GWC CY						
PpuXI	RG^GWC CY						
Psp5II	RG^GWC CY						
PspPPI	RG^GWC CY						
RsrII	CG^GWC CG	1119	781	347			
Rsr2I	CG^GWC CG	1119	781	347			
SanDI	GG^GWC CC						
SapI	GCTCTTCN^NNN	7260	4785	584	1296	1362	8870
SbfI	CC TGCA^GG	2591	1802	60	13	66	251
SdaI	CC TGCA^GG	2591	1802	60	13	66	251
SdiI	GGCCN NNN^NGGCC not available	2214	1634	28	18	54	121
SexAI	A^CCWGG T						
SfiI	GGCCN NNN^NGGCC	2214	1634	28	18	54	121
SgfI	GCG AT^CGC	108	62	207	39	29	178
SgrAI	CR^CCGG YG						
Sse232I	CG^CCGG CG not available	708	448	29	43	23	446
Sse1825I	GG^GWC CC not available						
Sse8387I	CC TGCA^GG	2591	1802	60	13	66	251
Sse8647I	AG^GWC CT not available						
VpaK32I	GCTCTTCN^NNN not available	7260	4785	584	1296	1362	8870

Six Cutters

Enzymes	Recognition Sequence	HsFna	MGC	Ec	Ce	Sc	At
NruI	TCG^CGA	830	607	1070	558	507	2422
SplI	C^GTAC G	701	449	498	263	549	1705
SnaBI	TAC^GTA	1080	621	435	165	885	2164
PvuI	CG AT^CG	842	512	1000	705	537	3078
MluI	A^CGCG T	1049	1019	976	295	337	1824
Bgl	GCCN NNN^NGGC	8827	6868	1333	469	750	2576
Ear I	CTCTTCN^NNN						
BsrGI	TGTACA	6683	4551	442	760	1583	5450
XmnI	GAANN^NNTTC	8401	5850	1167	1652	2911	12141
SaiI	G^TCGA C	1515	944	463	792	856	3616
BamHI	G^GATC C	6426	4305	438	1047	1238	6782
KpnI	G GTAC^C	4098	2755	442	305	1317	2992
EcoRI	G^AATT C	6538	4132	470	1346	2466	8244
XhoI	C^TCGA G	3651	2402	156	800	737	7092
EcoRV	GAT^ATC	3789	2435	1378	919	2289	8419
	AAA^TTT	7484	5167	1008	1049	3843	8078
DraI	TTT^AAA	8455	6243	967	494	3018	6778

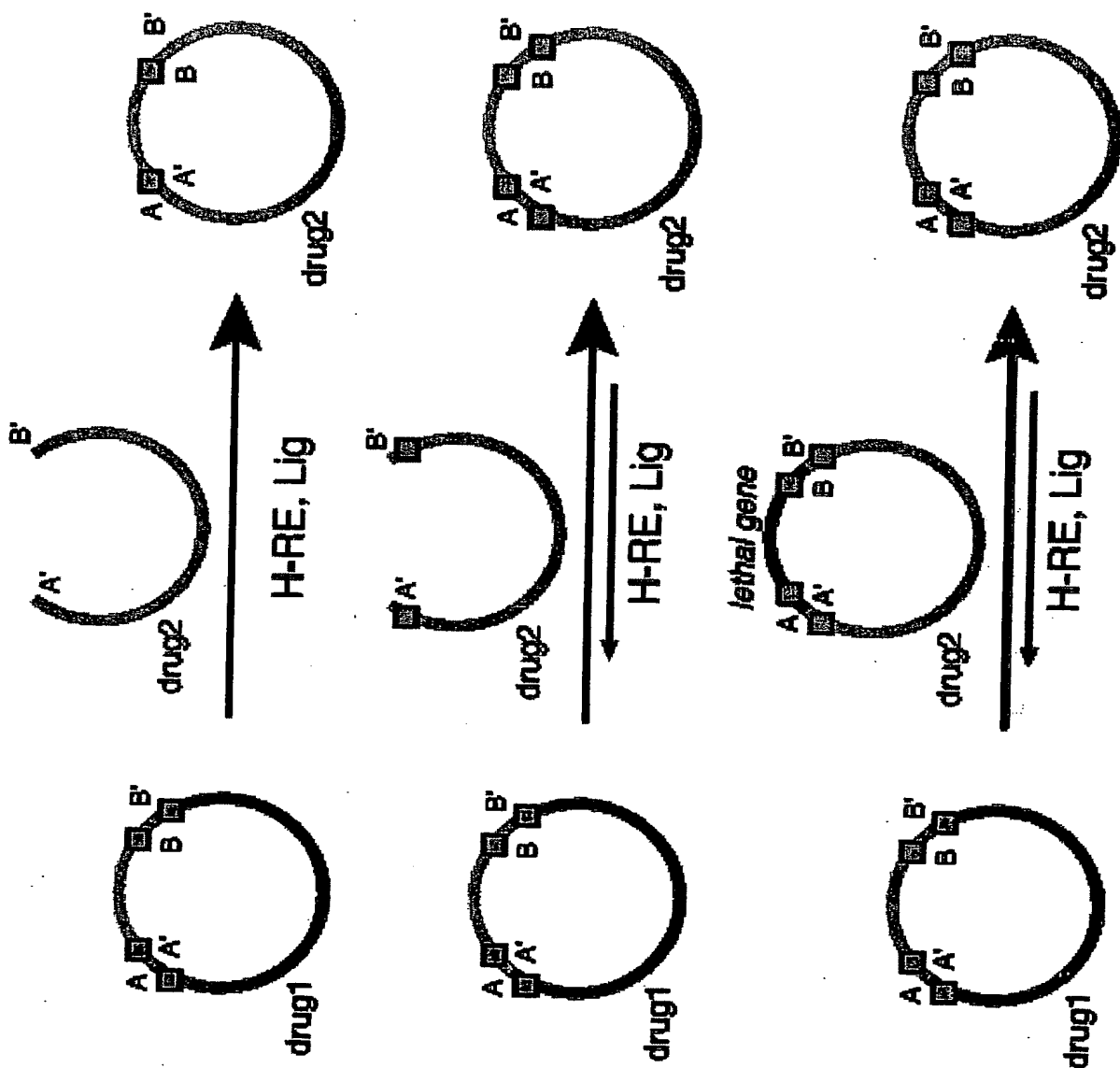
Fig 5

7+ Blunt Cutters

Enzymes	Recognition Sequence	HsFna	MGC	Ec	Ce	Sc	At
BstRZ246I	ATTT^AAAT	1204	648	55	38	379	317
BstSWI	ATTT^AAAT	1204	648	55	38	379	317
MspSWI	ATTT^AAAT	1204	648	55	38	379	317
MssI	GTTT^AAAC	297	173	71	9	152	490
PmeI	GTTT^AAAC	297	173	71	9	152	490
SmiI	ATTT^AAAT	1204	648	55	38	379	317
Swal	ATTT^AAAT	1204	648	55	38	379	317
SrfI	GCCC^GGGC	1433	887	40	11	11	30

Hs_Fna 15740 entries
 Mgc 12585 entries
 Ec 4290 entries
 Ce 3305 entries
 Sc 6360 entries
 At 27289 entries

Interrupted Palindromes



Sfi I

• How to make Sfi I “one way”

– Methylases

– Bgl I, not Sfi I sites, in *Acceptor* Vectors

G G C C N N N N^N G G C C
C C G G N^N N N N C C G G

G C C N N N N^N G G C
C G G N^N N N N C C G

– Lethal genes in stuffer fragments

PCR interrupted Palindrome Cloning Pathway

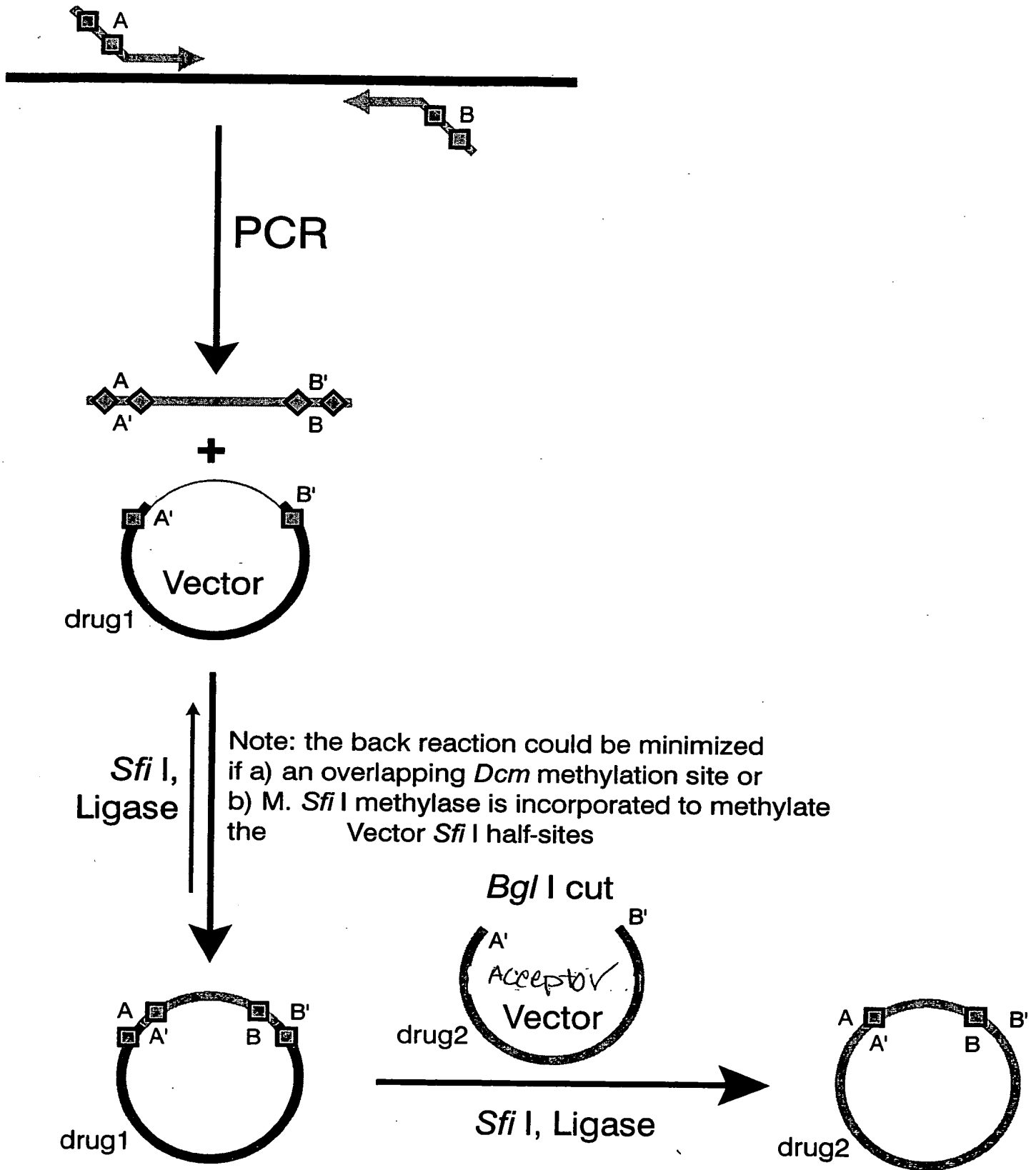
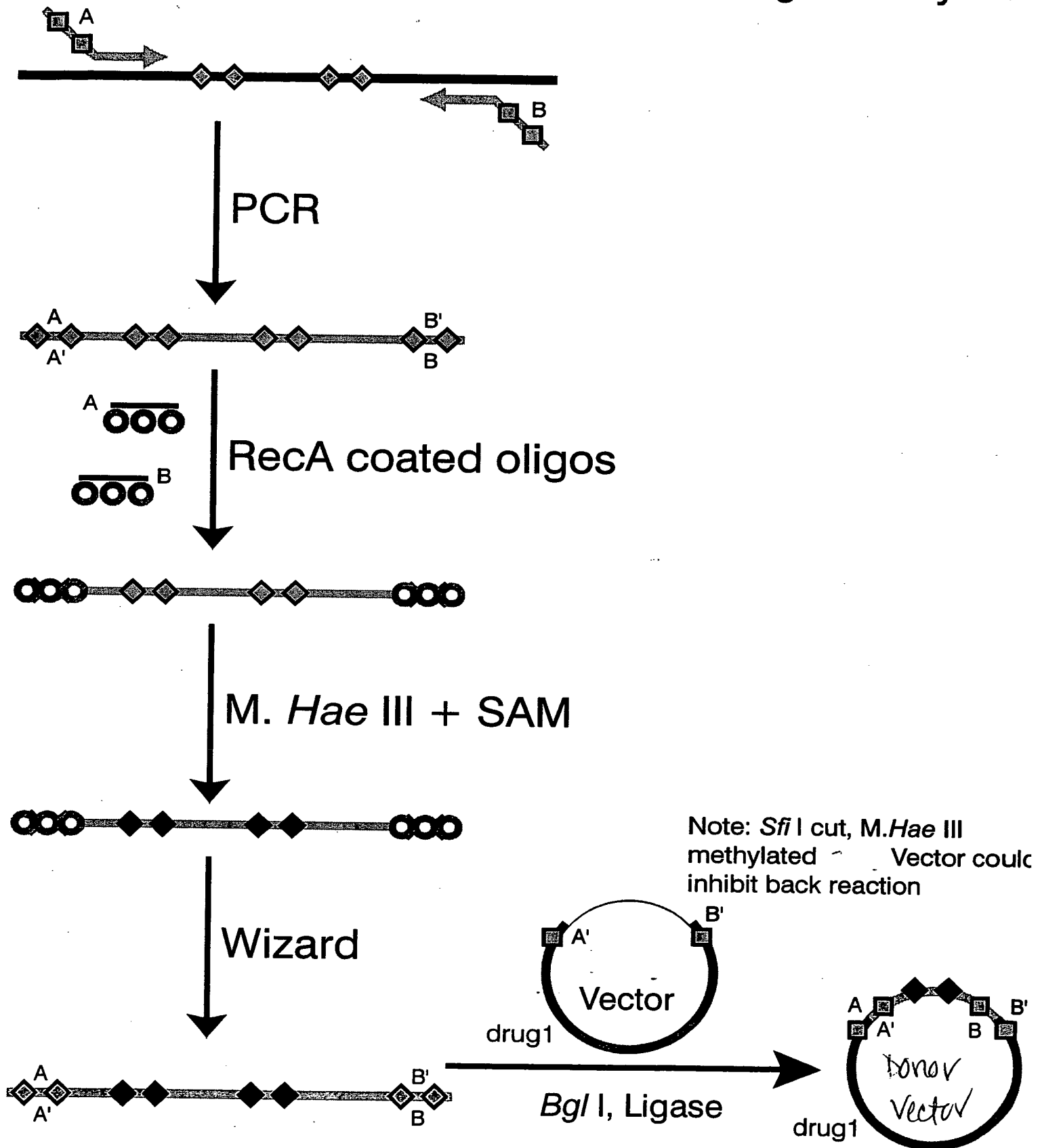


Fig8

PCR Interrupted Palindrome Cloning Pathway 1A



PCR Interrupted Palindrome Cloning Pathway 2A

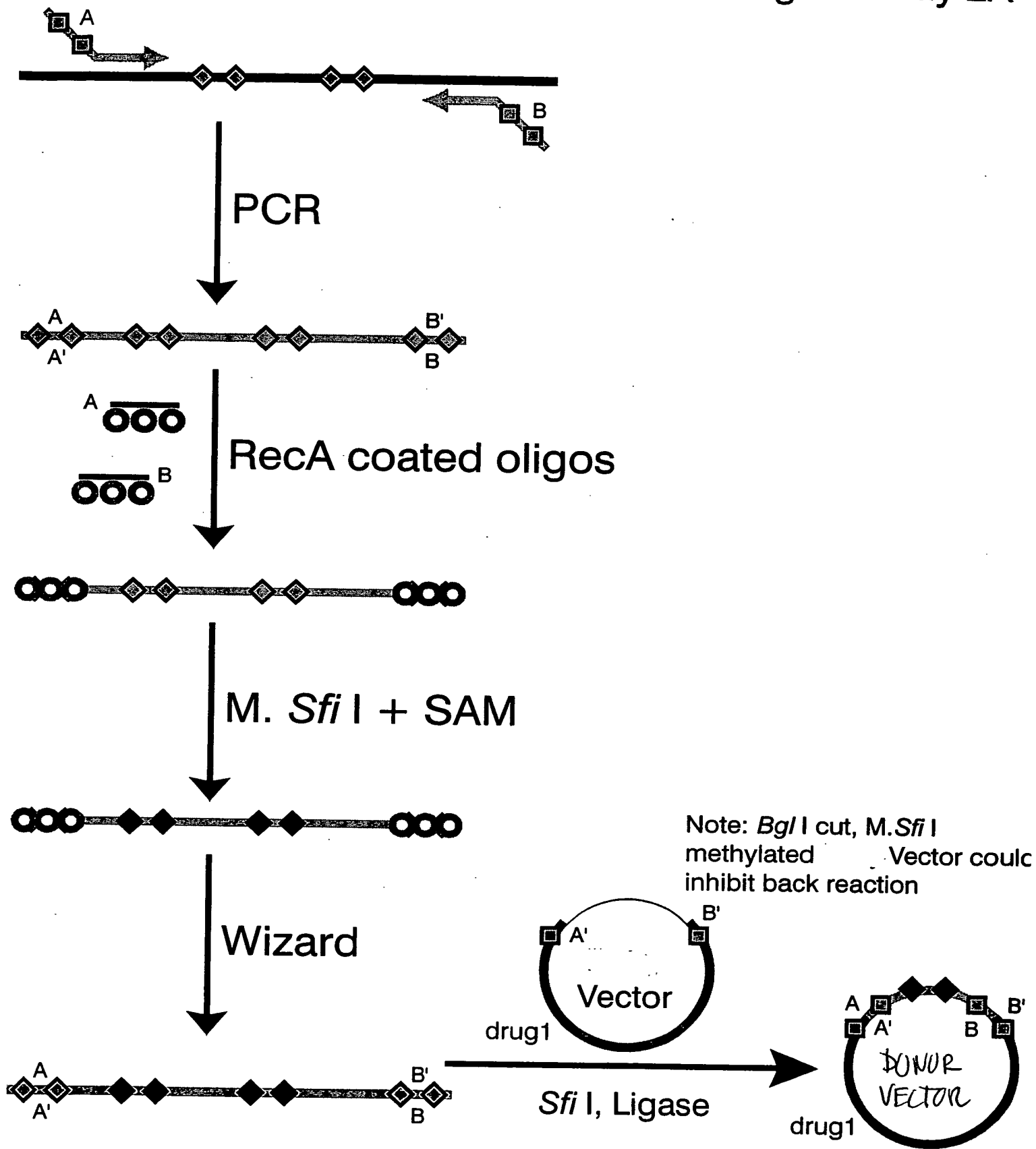
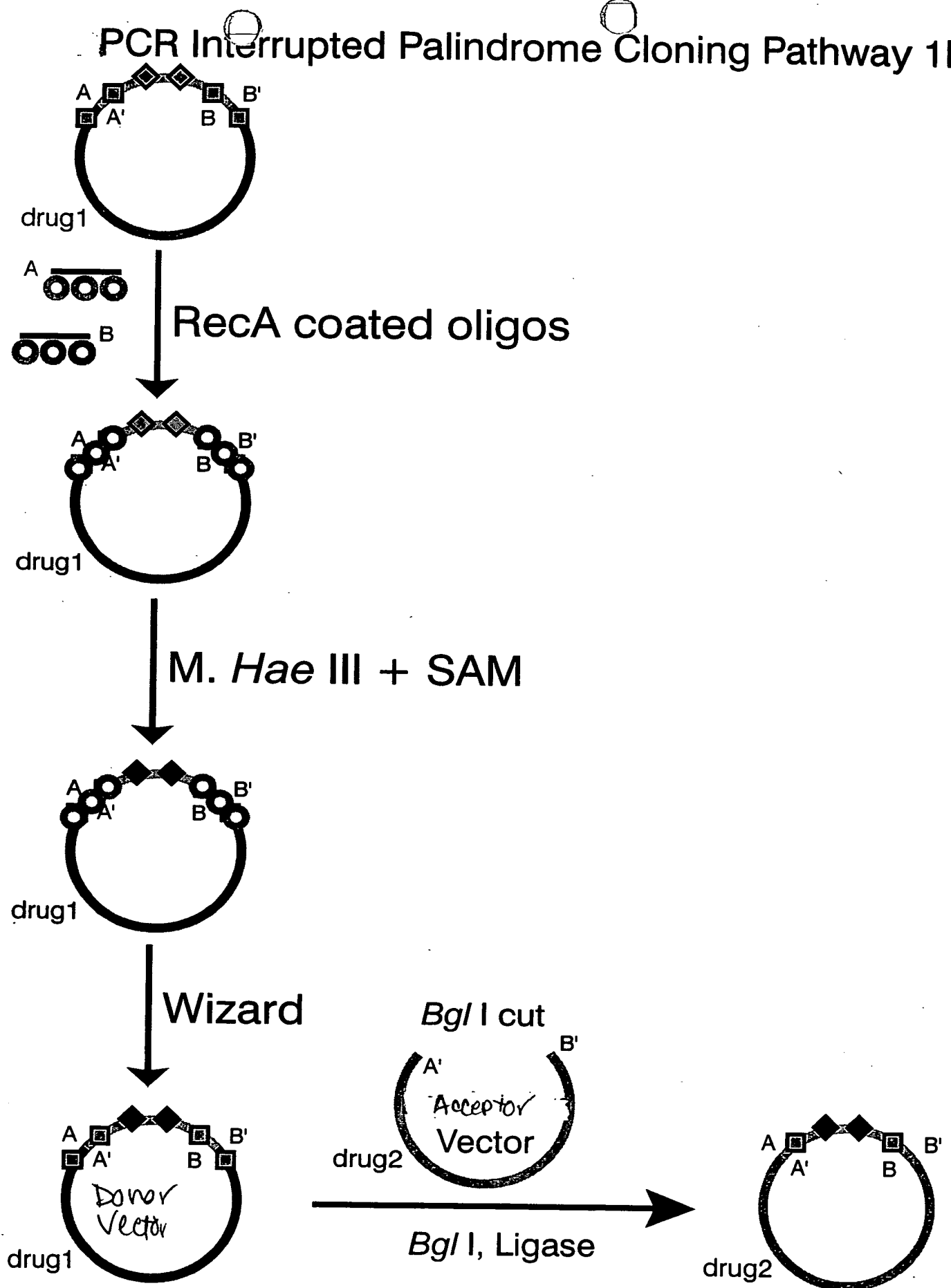
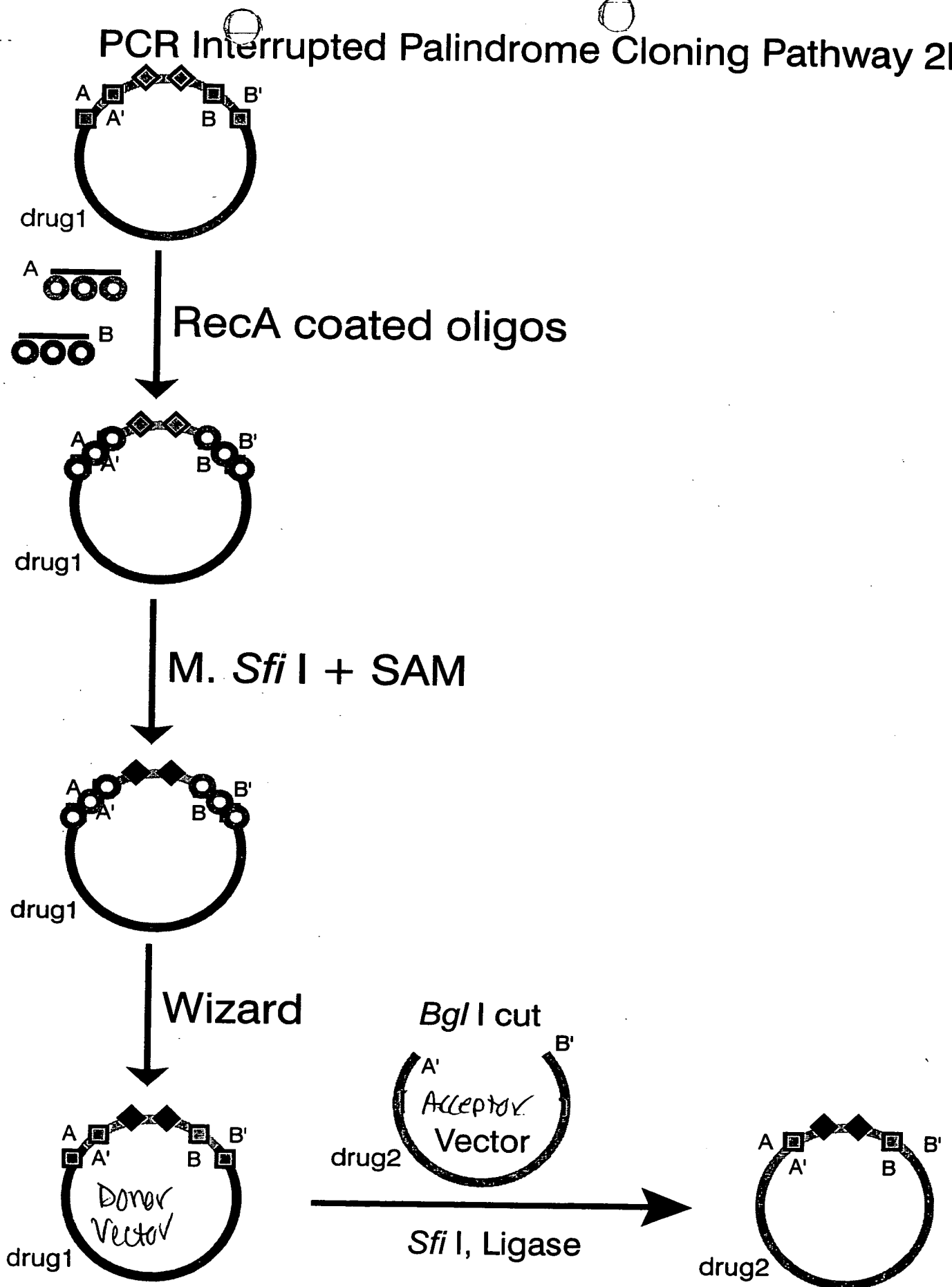


Fig 9B

PCR Interrupted Palindrome Cloning Pathway 1B



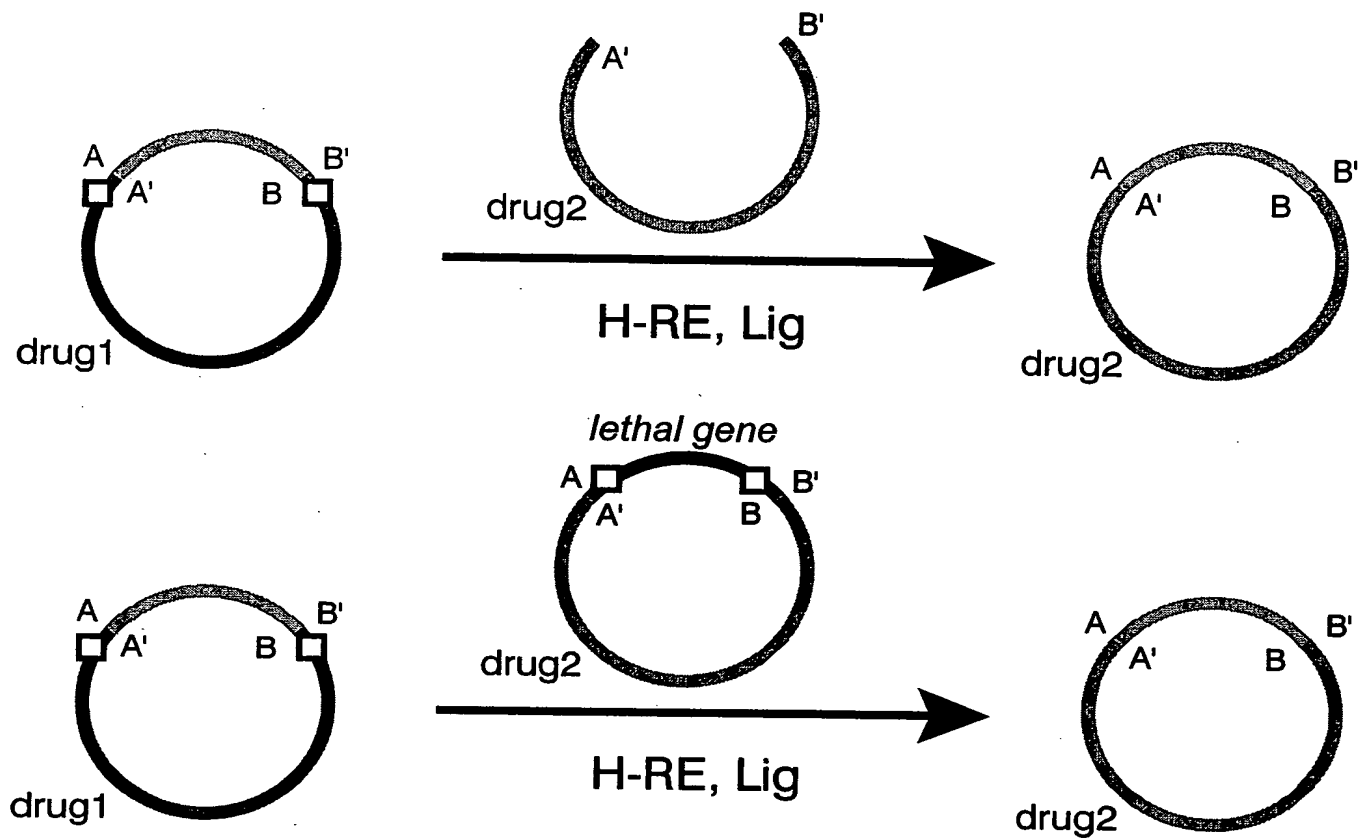
PCR Interrupted Palindrome Cloning Pathway 2B



REs that can make *Sfi* I one-way

3' 3b Overhang	Restriction Enzyme	Recognition Sequence
CNG	FnuI	G_GNC^C
CNG	PssI	RG_GNC^CY
CWG	Psp03I	G_GWC^C
GNC	BthCI	G_CNG^C
GSC	TauI	G_CSG^C
NNN	AlwNI	CAG_NNN^CTG
NNN	BglI	GCCN_NNN^NGGC
NNN	BsiYI	CCNN_NNN^NNGG
NNN	BstAPI	GCAN_NNN^NTGC
NNN	DraIII	CAC_NNN^GTG
NNN	MwoI	GCNN_NNN^NNGC
NNN	PfIMI	CCAN_NNN^NTGG
NNN	RleAI	CCCACANNNNNNN_NNN^
NNN	SfiI	GGCCN_NNN^NGGCC

Outside Cutters (Type IIS)



Sap I

How to make Sap I "one way"

- Methylases
- Orientation of sites in vector backbone in Donor Vector and in Acceptor Vector
- Lethal genes in stuffer fragments
- *Eco* I, not Sap I sites, in Acceptor Vectors

G C T C T C N[^] N N N
C G A G A A G N N N N[^]

C T C T C N[^] N N N
G A G A A G N N N N[^]

Key Advantage of Sap I

- Only three bases per exchange site left in Acceptor Vector

Two Enzyme Approach

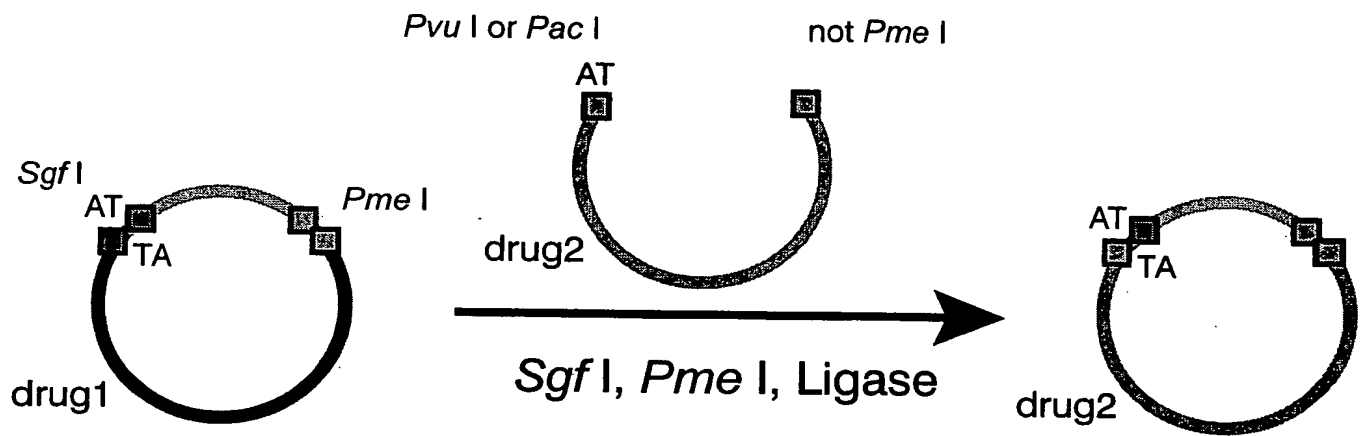
*Sgf*I – infrequent
cutter of human cDNAs, two base 3'
overhang

G C G A T[^]C G C
C G C[^]T A G C G

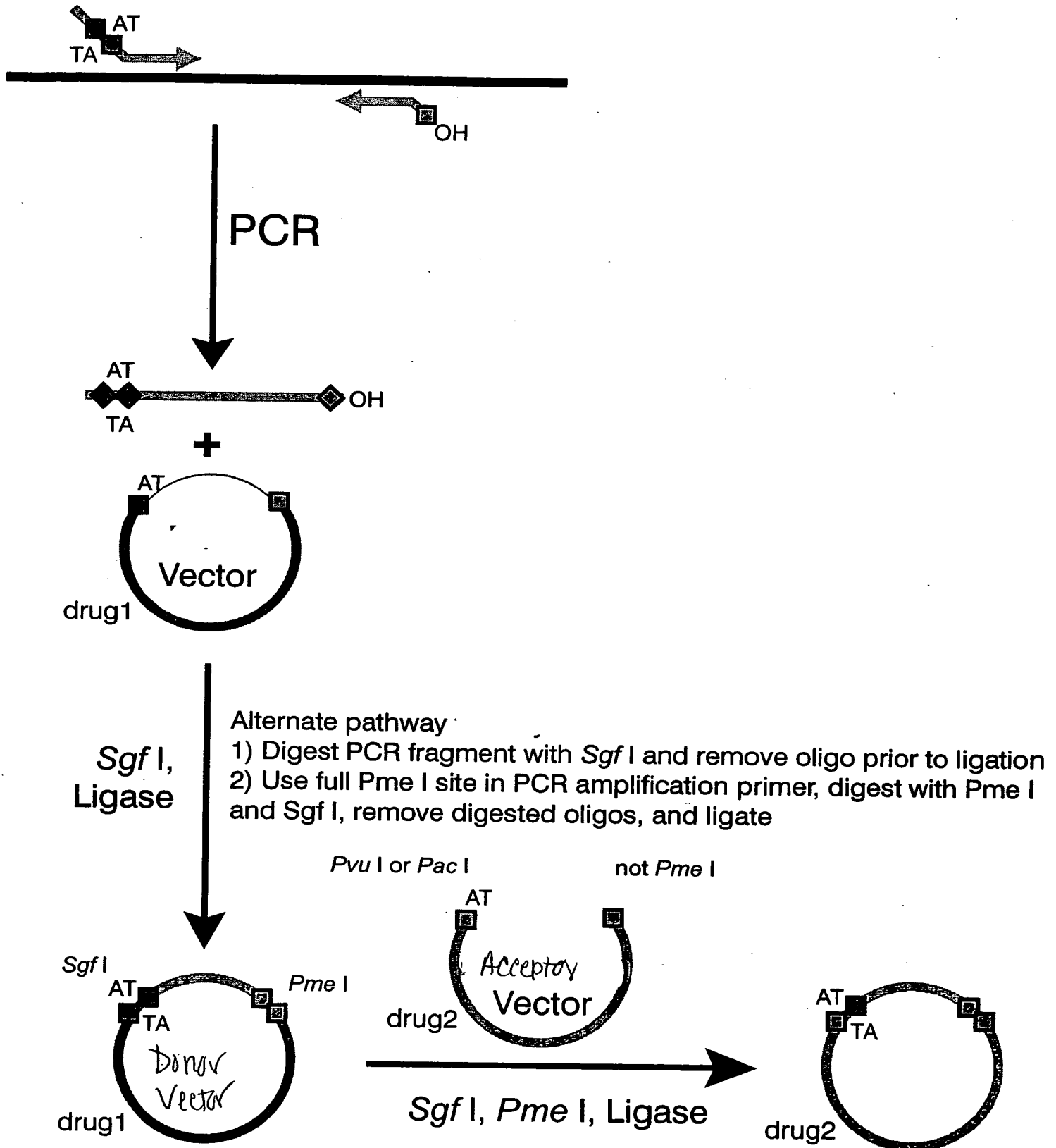
*Pme*I – infrequent
cutter, blunt end cutter

G T T T[^]A A A C
C A A A[^]T T T G

Two Enzyme Approach



Two Enzyme Cloning Pathway with PCR Entry



N-terminal SgfI site can allow N-terminal fusions OR *NO fusion*

NAAGGAGCGATCGCCATGg

--RBS- Kozak--

VAAGGAGCGATCGCCATG

KEQGlyAlaIleAlaMet

C-terminal *Pme* I site allows termination
(+1AA) or C-terminal fusions

NNNGTTTAAACN

XaaValTer

NNNGTTTATCN with *EcoRV*

XaaValTyr

NNNGTTTCCAN with *BalI*, etc.

XaaValSer

Coexpression Variation

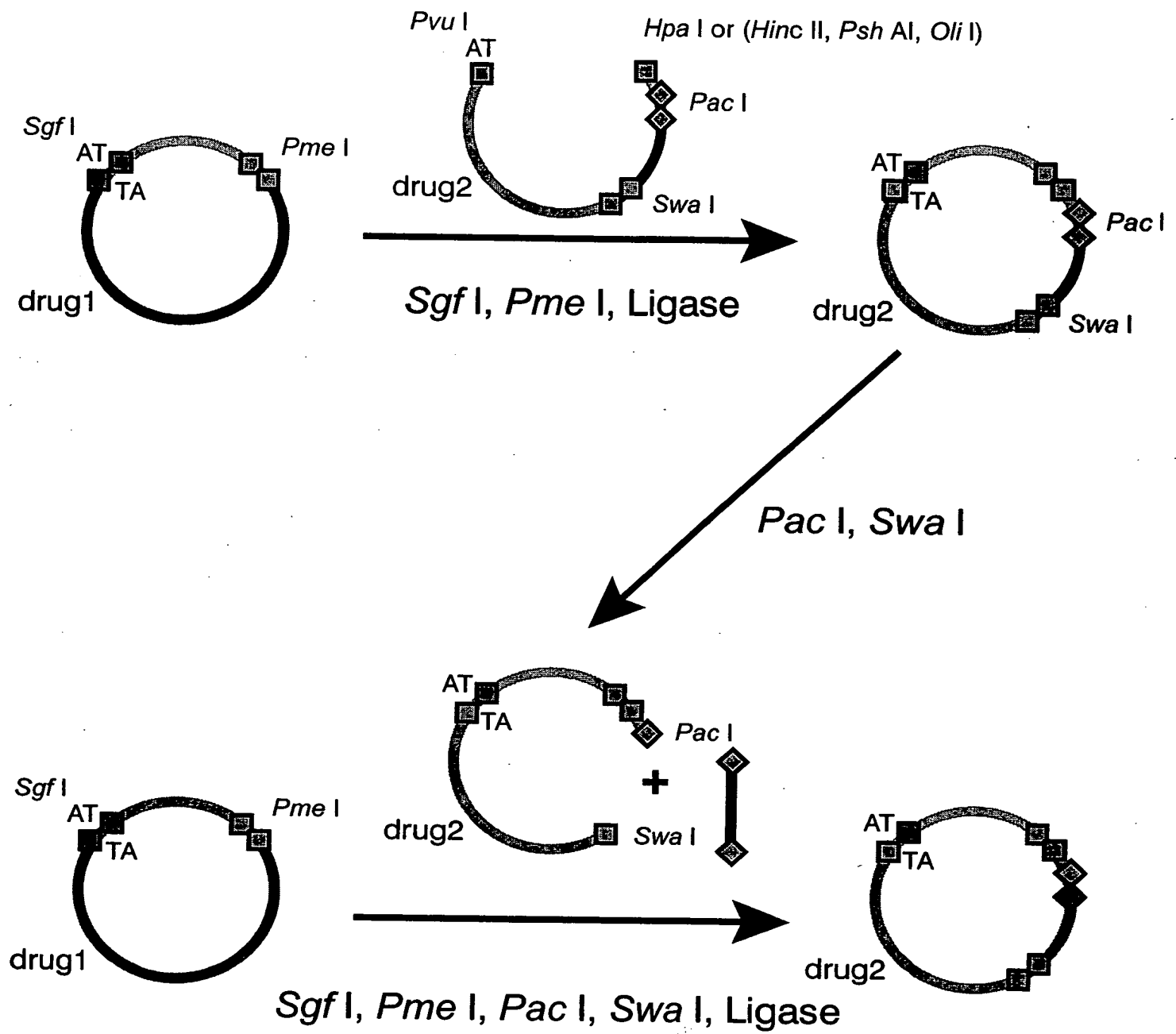


Fig 18

N-terminal *Pac* I--*Sgf* I fusion site

NAAGGATTAATCGCCATGg

KEQGlyLeuIleAlaMet

C-terminal *Pme* I--*Swa* I fusion site

NNNGTTTAAATN

XaaValTer

N-terminal *Pac* I--*Sgf* I fusion site

NAAGGATTAATCGCCATGg

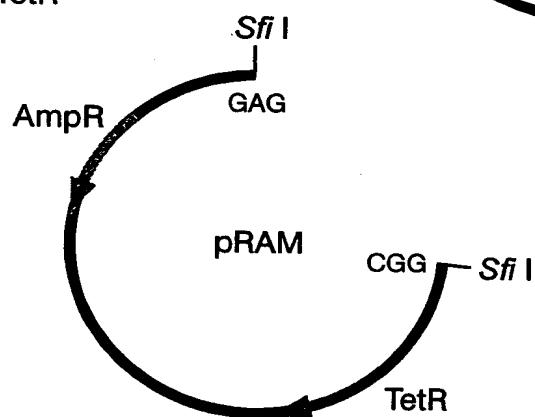
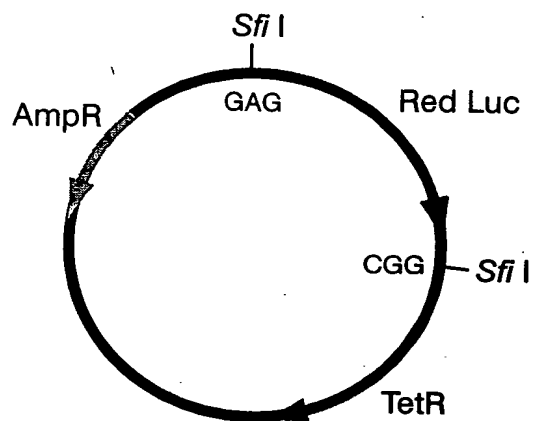
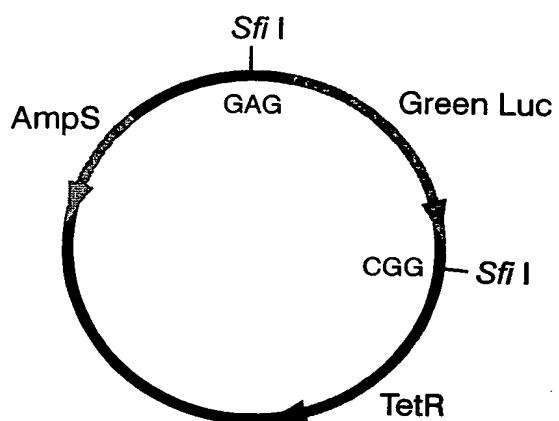
--RBS--

Kozak--

C-terminal *Pme* I--*Swa* I fusion site

NNNGTTTAAATN

XaaValTer



Sfi I cuts in ligase buffer, and cut ends religate with T4 DNA ligase

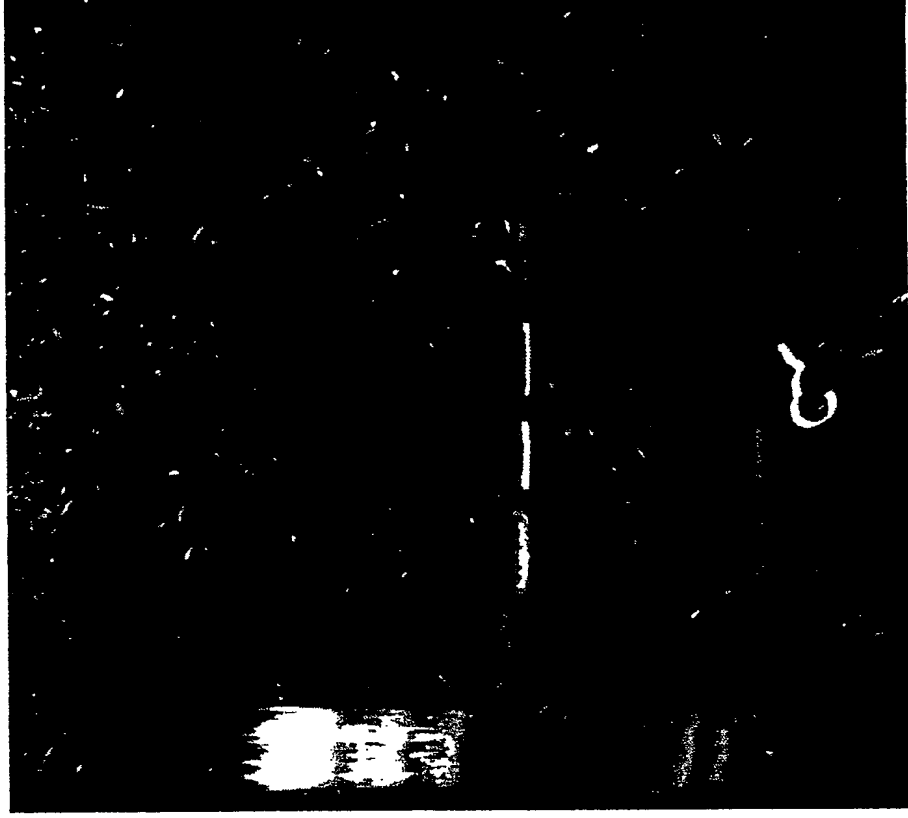
$i = N_0 M \times 10^3$ ends/ml for non-identical cohesive termini

$j = j\lambda (MW\lambda / MW)^{3/2}$ ends/ml

i = total concentration of DNA termini
 j = effective concentration of one end of a DNA molecule in the immediate neighborhood of the other end of the same molecule

Note: *Sfi* I cut at 50°C; ligase at 22°C

λ	no	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{1}$	$\frac{2}{1}$	$\frac{4}{1}$	i
H3	uc lig	4	2	1	1	1	i



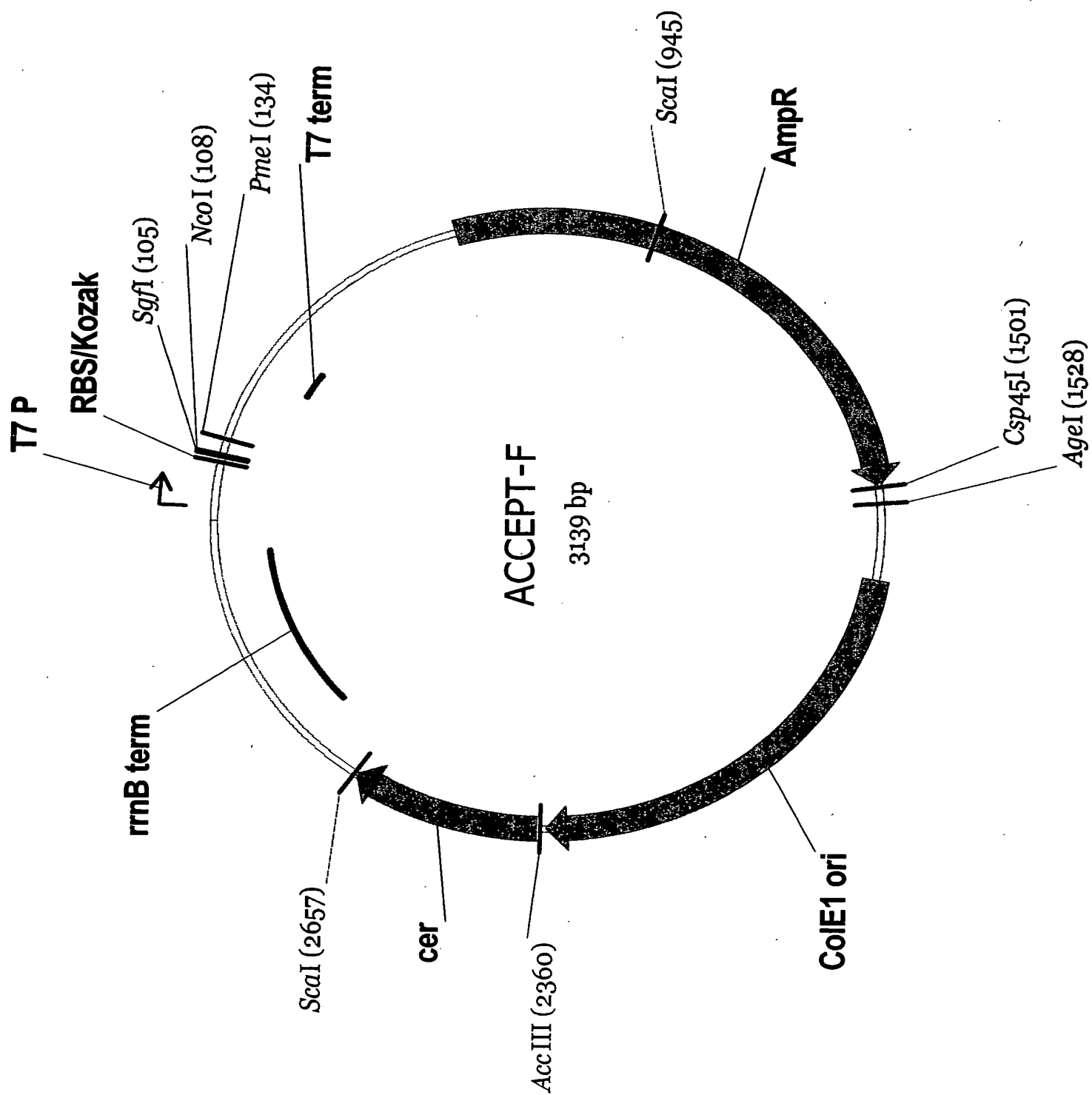
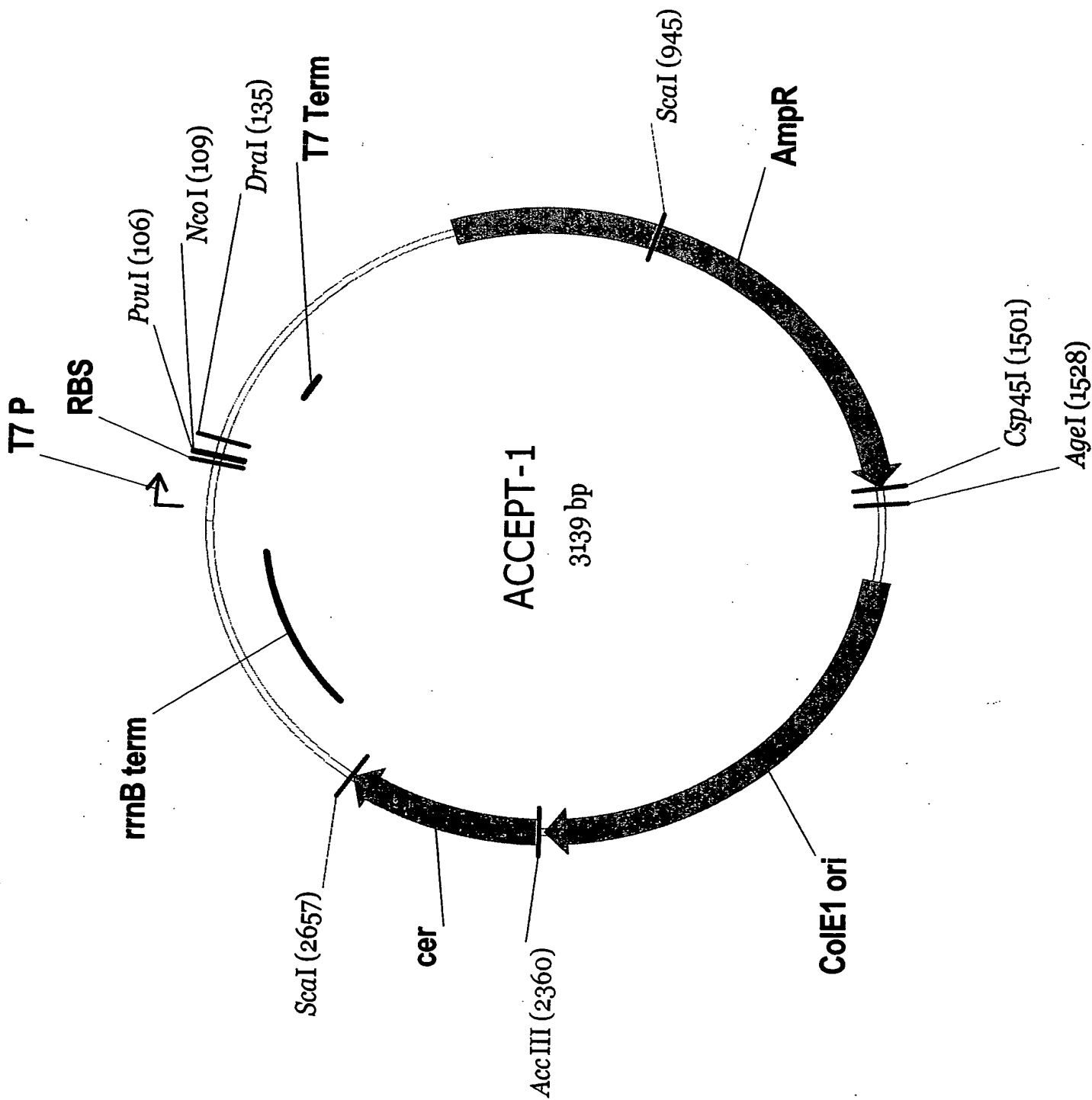
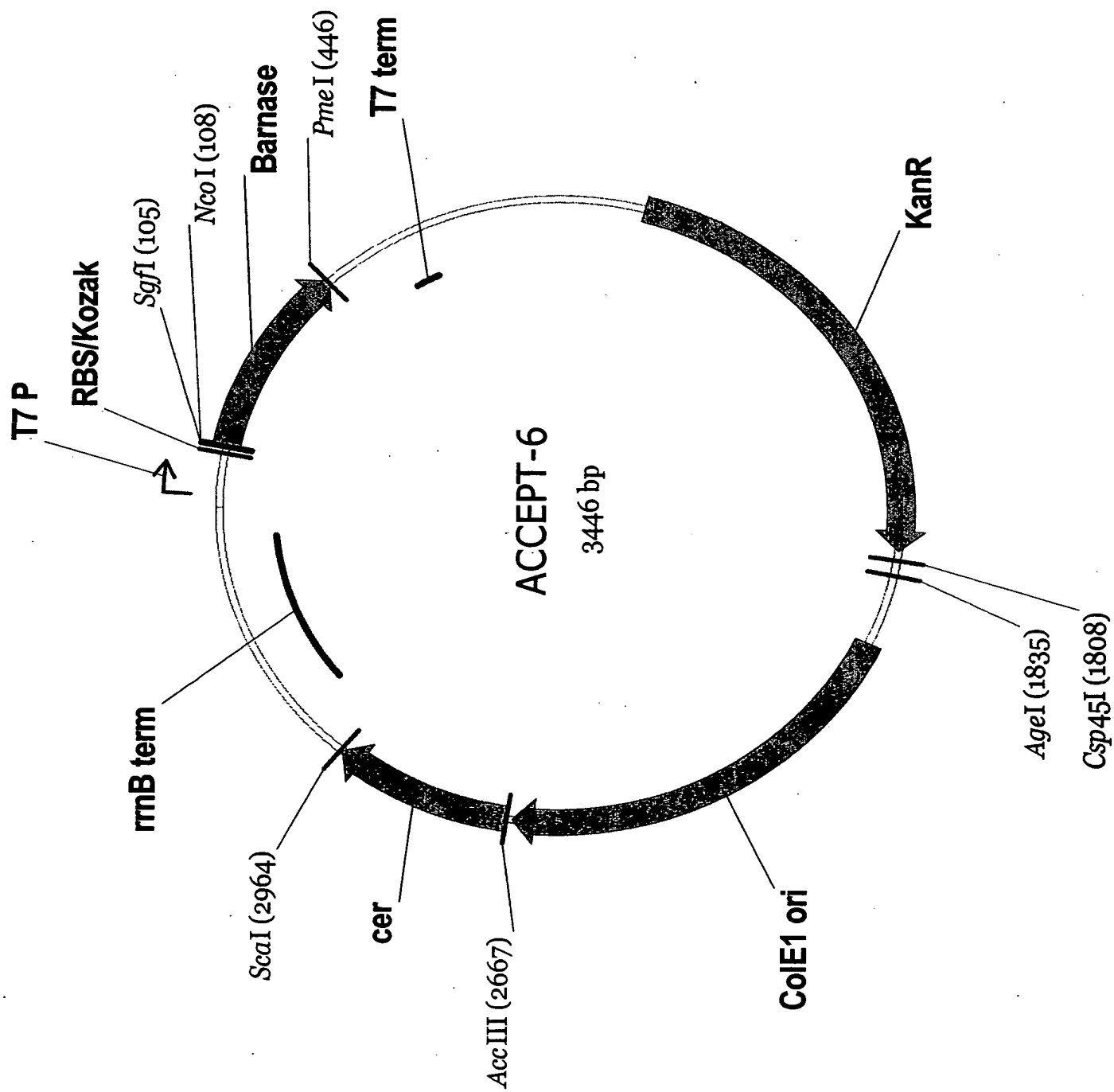
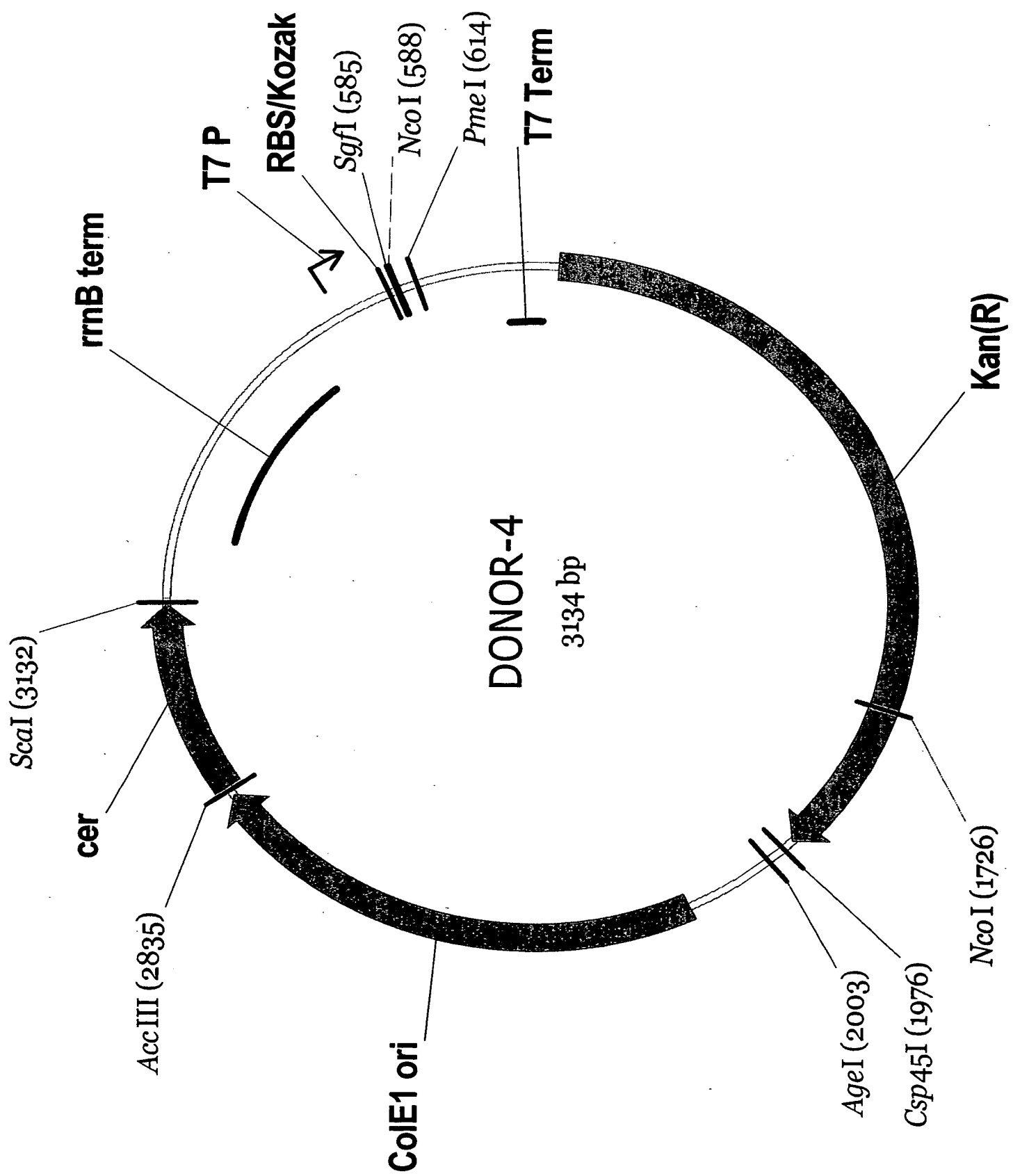
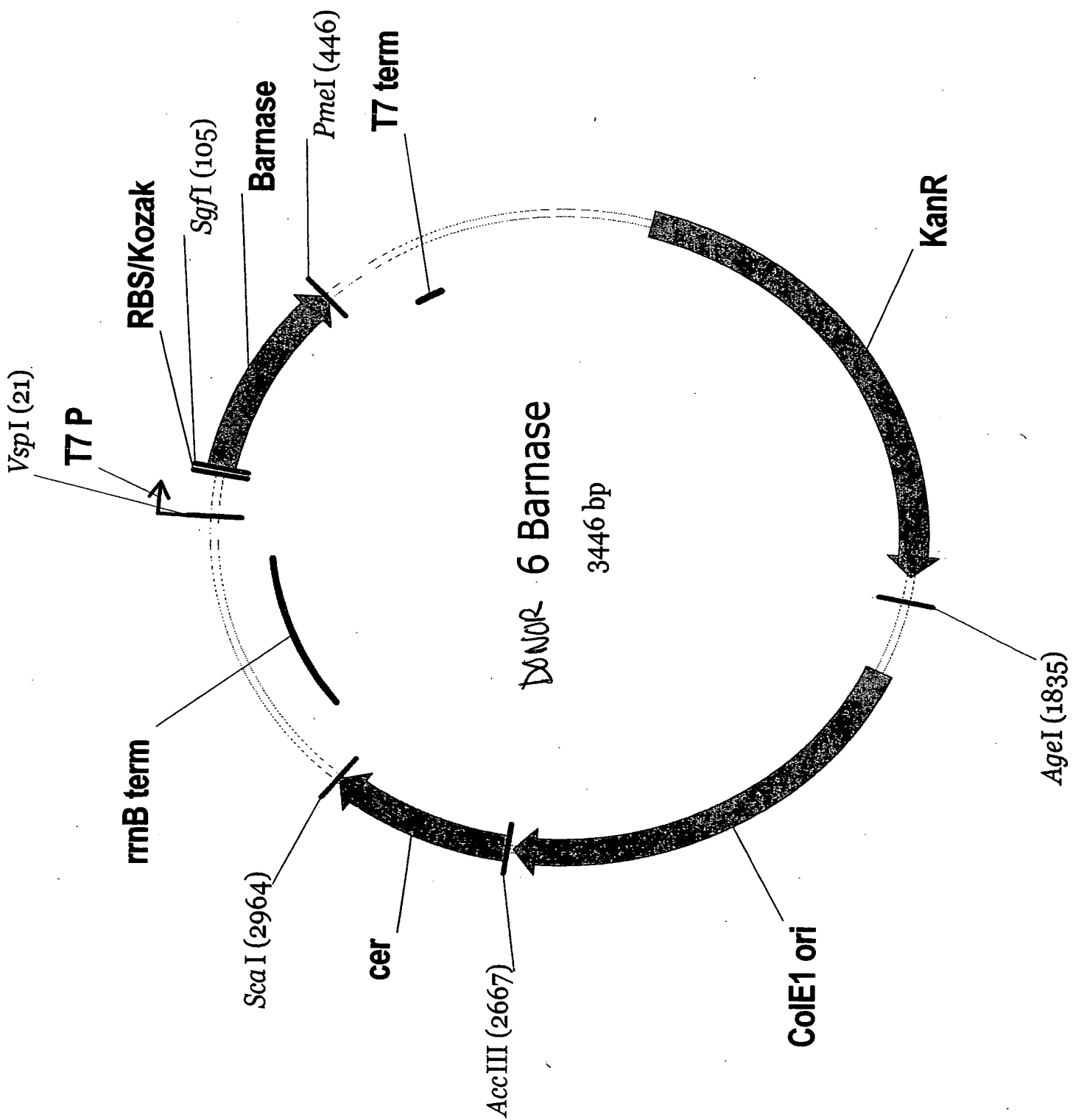


Fig 21A-E

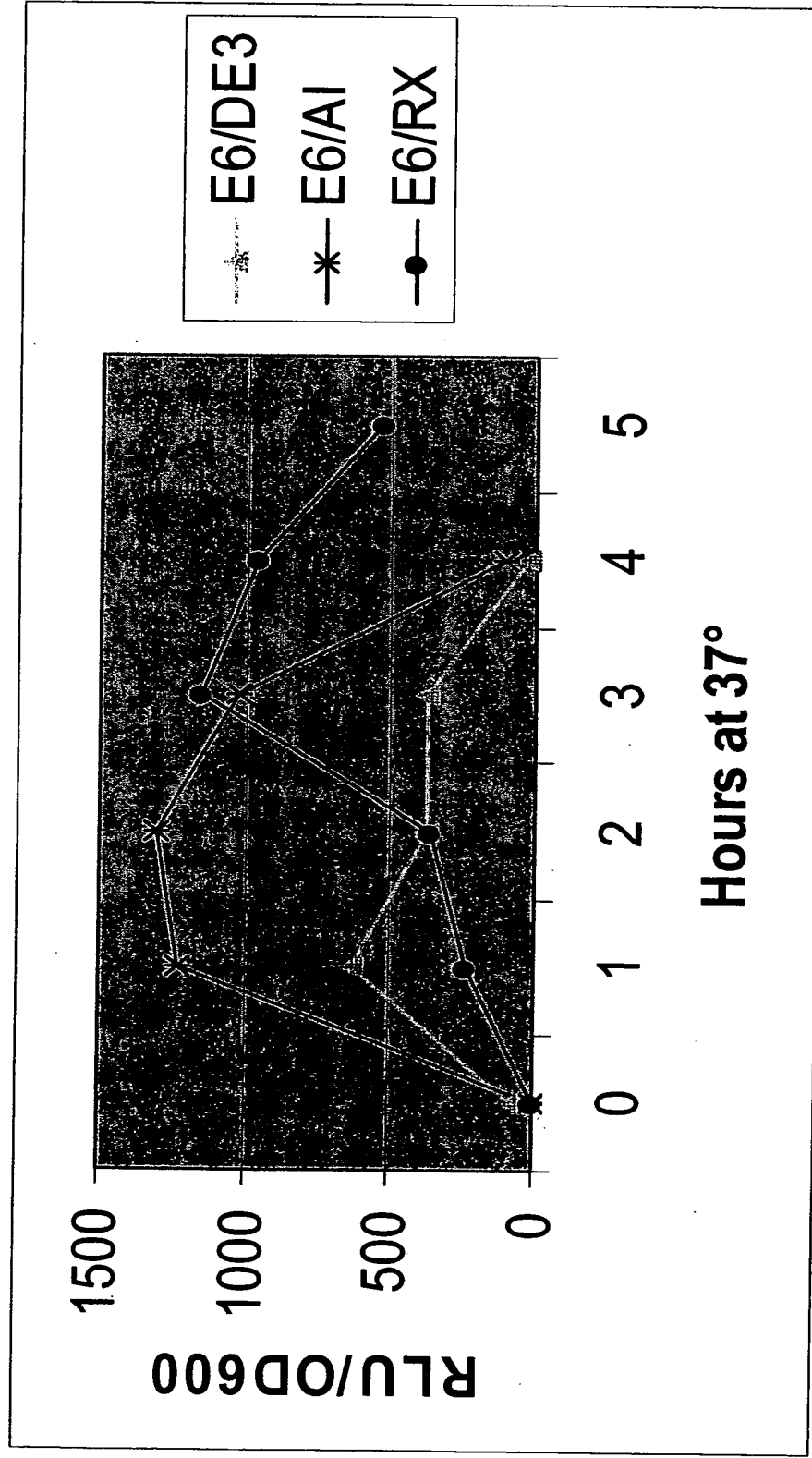




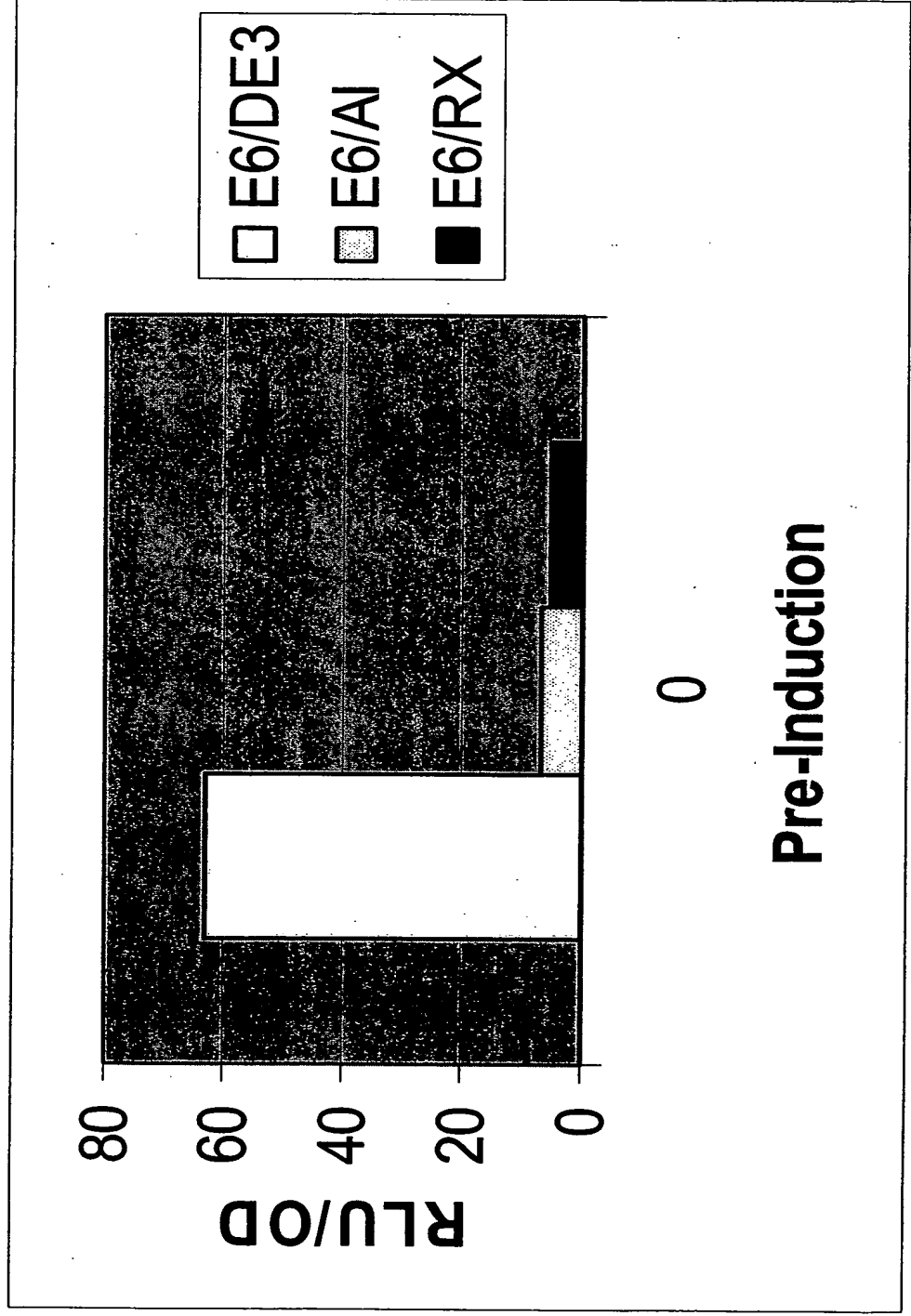




Luciferase Expression in 3 Hosts



Luciferase Expression in 3 Hosts at 25°C



Luciferase Expression in 3 Hosts at 25°C

